

# Nucleotide Sequence of the AIDS Virus, LAV

Simon Wain-Hobson,\* Pierre Sonigo,\*  
 Olivier Danos,† Stewart Cole,‡ and Marc Alizon§  
 \*Unité de Recombinaison et Expression Génétique  
 †Unité des Virus Oncogènes  
 ‡Groupe de Génie Génétique  
 §Unité d'Oncologie Virale  
 Institut Pasteur  
 25 et 28 rue du Dr. Roux  
 75724 Paris, Cedex 15, France

## Summary

The complete 9193-nucleotide sequence of the probable causative agent of AIDS, lymphadenopathy-associated virus (LAV), has been determined. The deduced genetic structure is unique: it shows, in addition to the retroviral gag, pol, and env genes, two novel open reading frames we call Q and F. Remarkably, Q is located between pol and env and F is half-encoded by the U3 element of the LTR. These data place LAV apart from the previously characterized family of human T cell leukemia/lymphoma viruses.

## Introduction

The recent onset of severe opportunistic infections among previously healthy male homosexuals has led to the characterization of the acquired immune deficiency syndrome (AIDS) (Gottlieb et al., 1981; Masur et al., 1981). The disease has spread dramatically, and new high-risk groups have been identified: patients receiving blood products, intravenous drug addicts, and individuals originating from Haiti and Central Africa (Piot et al., 1984). AIDS is a fatal disease, and there is at present no specific treatment. The causative agent was suspected to be of viral origin since the epidemiological pattern of AIDS was consistent with a transmissible disease, and cases had been reported after treatment involving ultrafiltered anti-hemophilia preparations (Daly and Scott, 1983). A decisive step in AIDS research was the discovery of a novel human retrovirus called lymphadenopathy-associated virus (LAV) (Barré-Sinoussi et al., 1983). The properties of the virus consistent with its etiological role in AIDS are: the recovery of many independent isolates from patients with AIDS or related diseases (Montagnier et al., 1984); high LAV seropositivity among these populations (Brun-Vézinet et al., 1984); a tropism and cytopathic effect in vitro for the helper/inducer T-lymphocyte subset T4 (Klatzmann et al., 1984), also found depleted in vivo.

Other groups have reported the isolation of human retroviruses, the human T cell leukemia/lymphoma/lymphotropic virus type III (HTLV-III) (Popovic et al., 1984) and the AIDS-associated retrovirus (ARV), which display biological and sero-epidemiological properties very similar to if not identical with those of LAV (Levy et al., 1984; Popovic et al., 1984; Schüpbach et al., 1984). Both LAV and HTLV-

III genomes have been molecularly cloned (Alizon et al., 1984; Hahn et al., 1984). Their restriction maps show remarkable agreement, including a Hind III restriction site polymorphism, bearing in mind the variability of this virus (Shaw et al., 1984) and confirming that these two viruses represent a single viral lineage.

In addition to its obvious diagnostic and therapeutic potential, the LAV DNA nucleotide sequence is essential to an understanding of the genetics and molecular biology of the virus and its classification among retroviruses. We report here the complete 9193-nucleotide sequence of the LAV genome established from cloned proviral DNA.

## Results

**DNA Sequence and Organization of the LAV Genome**  
 We have reported previously the molecular cloning of both cDNA and integrated proviral forms of LAV (Alizon et al., 1984). The recombinant phage clones were isolated from a genomic library of LAV-infected human T-lymphocyte DNA partially digested by Hind III. The insert of recombinant phage  $\lambda$ J19 was generated by Hind III cleavage within the R element of the long terminal repeat (LTR). Thus each extremity of the insert contains one part of the LTR. We have eliminated the possibility of clustered Hind III sites within R by sequencing part of an LAV cDNA clone, pLAV 75 (Alizon et al., 1984), corresponding to this region (data not shown). Thus the total sequence information of the LAV genome can be derived from the  $\lambda$ J19 clone.

Using the M13 shotgun cloning and dideoxy chain termination method (Sanger et al., 1977), we have determined the nucleotide sequence of  $\lambda$ J19 insert. The reconstructed viral genome with two copies of the R sequence is 9193 nucleotides long. The numbering system starts at the cap site (see below) of virion RNA (Figure 1).

The viral (+) strand contains the statutory retroviral genes encoding the core structural proteins (gag), reverse transcriptase (pol), and envelope protein (env), and two extra open reading frames (orf) that we call Q and F (Table 1). The genetic organization of LAV, 5LTR-gag-pol-Q-env-F3'LTR, is unique. Whereas in all replication-competent retroviruses pol and env genes overlap, in LAV they are separated by orf Q (192 amino acids) followed by four small (<100 triplets) orf. The orf F (206 amino acids) slightly overlaps the 3' end of env and is remarkable in that it is half-encoded by the U3 region of the LTR.

Such a structure clearly places LAV apart from previously sequenced retroviruses (Figure 2). The (-) strand is apparently noncoding. The additional Hind III site of the LAV clone  $\lambda$ J81 (with respect to  $\lambda$ J19) maps to the apparently noncoding region between Q and env (positions 5166-5745). Starting at position 5501 is a sequence (AAGCCT) that differs by a single base (underlined) from the Hind III recognition sequence. It is anticipated that many of the restriction site polymorphisms between different isolates will map to this region.

Kun I

## Nucleotide Sequence of the AIDS Virus, LAV

1

L'ESPRESSO

Bgl II

Figure 1. Complete DNA Sequence of Viral Genome (LAV-18)

The sequence was reconstructed from the sequence of phage L19 insert. The numbering starts at the cap site, which was located experimentally (see above). Important genetic elements, major open reading frames, and their predicted products are indicated together with the Hind III cloning sites. The potential glycosylation sites in the env gene are overlined. The NH<sub>2</sub>-terminal sequence of p25<sup>env</sup> determined by protein microsequencing is boxed (Genetic Systems, personal communication).

Each nucleotide was sequenced on average 5.3 times; 85% of the sequences was determined on both strands and the remaining 15% at least twice from independent clones. The base composition is T, 22.2%; C, 17.6%; A, 35.8%; G, 24.2%; G + C, 42%. The dinucleotide CpG is greatly under-represented (0.9%) as is common among eukaryotic sequences (Bird, 1980).

The LTR

The organization of a reconstructed LTR and viral flanking elements are shown schematically in Figure 3. The LTR is 638 bp long and displays usual features (Chen and Barker, 1984); it is bounded by an inverted repeat (5'ACTG) including the conserved TG dinucleotide (Termin, 1981); adjacent to 5' LTR is the tRNA primer binding site (PBS), complementary to tRNA<sup>Ura</sup> (Raba et al., 1979); adjacent to 3' LTR is a perfect 15 bp polypurine tract. The other three

polypurine tracts observed between nucleotides 8200-8800 are not followed by a sequence that is complementary to that just preceding the PBS.

The limits of U5, R, and U3 elements were determined as follows. U5 is located between PBS and the polyadenylation site established from the sequence of the 3' end of oligo(dT)-primed LAV cDNA (Alizon et al., 1984). Thus U5 is 84 bp long. The length of R+U5 was determined by synthesizing tRNA-primed LAV cDNA. After alkaline hydroly-

**Table 1. Locations and Sizes of Viral Open Reading Frames**

| orf   | 1 <sup>st</sup> Triplet | Met   | Stop  | No. Amino Acids | M, Calc.  |
|-------|-------------------------|-------|-------|-----------------|-----------|
| gag   | 312                     | 336   | 1 836 | 500             | 55,841    |
| pol   | 1,631                   | 1,934 | 4,640 | (1,003)         | (113,629) |
| orf Q | 4,554                   | 4,587 | 5,163 | 192             | 22,487    |
| env   | 5,746                   | 5,767 | 8,350 | 861             | 97,376    |
| orf E | 8,924                   | 8,354 | 8,972 | 206             | 23,316    |

The nucleotide coordinates refer to the first base of the first triplet (1<sup>st</sup> triplet), of the first methionine (initiation) codon (Met) and of the stop codon (Stop). The numbers of amino acids and molecular weights are those calculated for unmodified precursor products starting at the first methionine through to the end, with the exception of pol., where the size and M<sub>r</sub> refer to that of the whole ori.

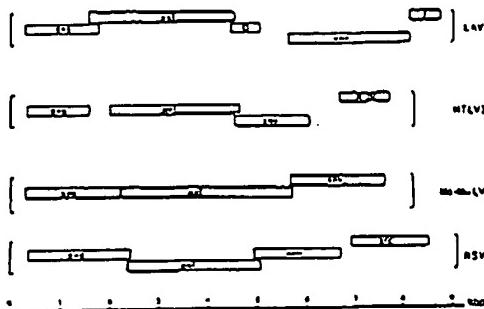


Figure 2. Comparison of the Genome Organization of LAV with Those of Human T Cell Leukemia/Lymphoma Virus Type I (HTLV-I) (Seiki et al., 1983), Moloney Murine Leukemia Virus (MoMuLV) (Shinnick et al., 1981), and Rous Sarcoma Virus (RSV) (Schwartz et al., 1983). The positions and sizes of viral genes are drawn to scale (open boxes) and the viral genome (RNA forms) are delimited by brackets.

sis of the primer, R+U5 was found to be  $181 \pm 1$  bp (Figure 4). Thus R is 97 bp long and the cap site at its 5' end can be located. Finally, U3 is 456 bp long. The LAV LTR also contains characteristic regulatory elements: a polyadenylation signal sequence AATAAA 19 bp from the R-U5 junction, and the sequence ATATAAG, which is very likely the TATA box, 22 bp 5' of the cap site. There are no long direct repeats within the LTR. Interestingly, the LAV LTR shows some similarities to that of the mouse mammary tumor virus (MMTV) (Donehower et al., 1981). They both use tRNA<sup>Val</sup> as a primer for (-) strand synthesis, whereas all other exogenous mammalian retroviruses known to date use tRNA<sup>Pro</sup> (Chen and Barker, 1984). They possess very similar polypurine tracts; that of LAV is AAAAGAAAAGGGGG while that of MMTV is AAAAAAGAAAAAAGGGGG. It is probable that the viral (+) strand synthesis is discontinuous since the polypurine tract flanking the U3 element of the 3'LTR is found exactly duplicated in the 3' end of orf pol, at 4331-4346. In addition, MMTV and LAV are exceptional in that the U3 element can encode an orf. In the case of MMTV, U3 contains the whole orf while, in LAV, U3 contains 110 codons of the 3' half of orf F.

#### Viral Proteins

##### gag

Near the 5' extremity of the gag orf is a "typical" initiation codon (Kozak, 1984) (position 336), which is not only the first in the gag orf, but the first from the cap site. The precursor protein is 500 amino acids long. The calculated  $M_r$  of 55,841 agrees with the 55 kd gag precursor polypeptide (Luc Montagnier, unpublished results). The N-terminal amino acid sequence of the major core protein p25, obtained by microsequencing (Genetic Systems, personal communication), matches perfectly with the translated nucleotide sequence starting from position 732 (see Figure 1). This formally makes the link between the cloned LAV genome and the immunologically characterized LAV p25 protein. The protein encoded 5' of the p25 coding sequence is rather hydrophilic. Its calculated  $M_r$  of 14,865 is consistent with that of the gag protein p18. The 3' part of the gag region probably codes for the retroviral nucleic acid binding protein (NBP). Indeed, as in HTLV-I (Seiki et

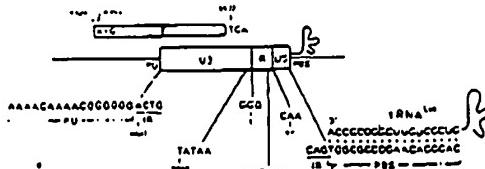


Figure 3. Schematic Representation of the LAV Long Terminal Repeat (LTR)

The LTR was reconstructed from the sequence of LAV by juxtaposing the sequences adjacent to the Hind III cloning sites. Sequencing of oligo(dT)-primed LAV DNA clone pLAV75 (Alizon et al., 1984) rules out the possibility of clustered Hind III sites in the R region of LAV. LTR are limited by an inverted repeat sequence (IR). Both of the viral elements flanking the LTR have been represented as tRNA primer binding site (PBS) for 5' LTR and polyuridine tract (PU) for 3' LTR. Also indicated are a putative TATA box, the cap site, polyadenylation signal (AATAAA), and polyadenylation site (CAA). The location of the open reading frame F (648 nucleotides) is shown above the LTR scheme.

al., 1983) and RSV (Schwartz et al., 1983), the motif Cys-X<sub>2</sub>-Cys-X<sub>n-9</sub>-Cys common to all NBP (Oroszlan et al., 1984) is found duplicated (nucleotides 1509 and 1572 in LAV sequence). Consistent with its function the putative NBP is extremely basic (17% Arg + Lys).

##### pol

The reverse transcriptase gene can encode a protein of up to 1003 amino acids (calculated  $M_r \approx 113,629$ ). Since the first methionine codon is 92 triplets from the origin of the open reading frame, it is possible that the protein is translated from a spliced messenger RNA, giving a gag-pol polyprotein precursor.

The pol coding region is the only one in which significant homology has been found with other retroviral protein sequences, three domains of homology being apparent. The first is a very short region of 17 amino acids (starting at 1856). Homologous regions are located within the p15 gag<sup>RSV</sup> protease (Dittmar and Moelling, 1978) and a polypeptide encoded by an open reading frame located between gag and pol of HTLV-I (Figure 5) (Schwartz et al., 1983; Seiki et al., 1983). This first domain could thus correspond to a conserved sequence in viral proteases. Its different locations within the three genomes may not be significant since retroviruses, by splicing or other mechanisms, express a gag-pol polyprotein precursor (Schwartz et al., 1983; Seiki et al., 1983). The second and most extensive region of homology (starting at 2048) probably represents the core sequence of the reverse transcriptase. Over a region of 250 amino acids, with only minimal insertions or deletions, LAV shows 38% amino acid identity with RSV, 25% with HTLV-I, and 21% with MoMuLV (Shinnick et al., 1981) while HTLV-I and RSV show 38% identity in the same region. A third homologous region is situated at the 3' end of the pol reading frame and corresponds to part of the pp32 peptide of RSV that has exonuclease activity (Misra et al., 1982). Once again, there is greater homology with the corresponding RSV sequence than with HTLV-I.

##### env

The env open reading frame has a possible initiator methionine codon very near the beginning (eighth triplet).

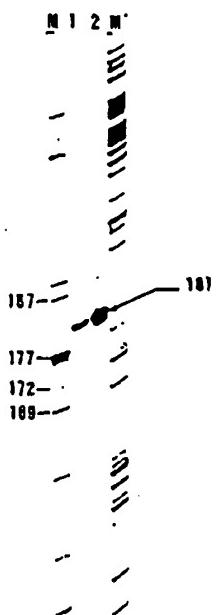


Figure 4. Synthesis of RNA-Primed LAV cDNA for R+US (Strong-Stop cDNA)

Lanes 1 and 2 show two different quantities of cDNA while lanes M and M' represent markers. The strong-stop cDNA is 181 bases long with a second, less intense band at 160. The error of estimation is  $\pm 1$  bp. This maps the major cap site to the second G residue of the sequence CTGGGTCT within the LTR, 24 nucleotides downstream of the TATA box. This guanosine residue is taken as the first base in the nucleotide sequence shown in Figure 1.

If so, the molecular weight of the presumed env precursor protein (861 amino acids, M, calc = 97,376) is consistent with the known size of the LAV glycoprotein (110 kd and 90 kd after glycosidase treatment; Luc Montagnier, unpublished). There are 32 potential N-glycosylation sites (Asn-X-Ser/Thr), which are overlined in Figure 1. An interesting feature of env is the very high number of Trp residues at both ends of the protein. There are three hydrophobic regions, characteristic of the retroviral envelope proteins (Seiki et al., 1983), corresponding to a signal peptide (encoded by nucleotides 5815–5850 bp), a second region (7315–7350 bp), and a transmembrane segment (7831–7896 bp). The second hydrophobic region (7315–7350 bp) is preceded by a stretch rich in Arg + Lys. It is possible that this represents a site of proteolytic cleavage, which, by analogy with other retroviral proteins, would give an external envelope polypeptide and a membrane-associated protein (Seiki et al., 1983; Kiyokawa et al., 1984). A striking feature of the LAV envelope protein sequence is that the region following the transmembrane segment is of unusual length (150 residues). The env protein shows no

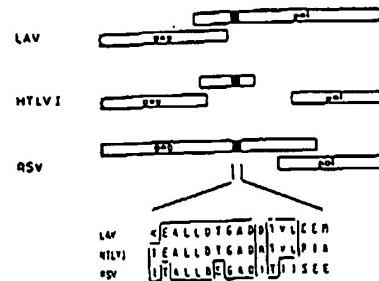


Figure 5. Location of a Short Stretch of Homology in the gag-pol Region of the LAV, HTLV-I (Seiki et al., 1983) and RSV (Schwartz et al., 1983) Genomes

Conserved amino acids are boxed. Homologous region is shown by the solid bar in the schema. Each virus is organized differently in this region but the sequence in the RSV genome maps to p15<sup>040</sup>, which has a protease-associated function.

homology to any sequence in protein data banks. The small amino acid motif common to the transmembrane proteins of all leukemogenic retroviruses (Cianciolo et al., 1984) is not present in LAV env.

#### Q and F

The location of orf Q is without precedent in the structure of retroviruses. Orf F is unique in that it is half-encoded by the U3 element of the LTR. Both orf have strong initiator codons (Kozak, 1984) near their 5' ends and can encode proteins of 192 amino acids (M, calc = 22,487) and 206 amino acids (M, calc = 23,316), respectively. Both putative proteins are hydrophilic (pQ 49% polar, 15.1% Arg + Lys; pF 46% polar, 11% Arg + Lys) and are therefore unlikely to be associated directly with membrane. The function for the putative proteins pQ and pF cannot be predicted, as no homology was found by screening protein sequence data banks. Between orf F and the pX protein of HTLV-I there is no detectable homology. Furthermore, their hydrophobicity/hydrophilicity profiles are completely different. It is known that retroviruses can transduce cellular genes—notably proto-oncogenes (Weinberg, 1982). We suggest that orfs Q and F represent exogenous genetic material and not some vestige of cellular DNA because LAV DNA does not hybridize to the human genome under stringent conditions (Alizon et al., 1984), and their codon usage is comparable to that of the gag, pol, and env genes (data not shown).

#### Relationship to Other Retroviruses

Although LAV is both morphologically and biochemically (Barré-Sinoussi et al., 1983) distinct to HTLV-I and -II, it remained possible that its genome was organized in a similar manner. The characteristic features of HTLV-I and -II genomes, which they share with the more distantly related bovine leukemia virus (BLV) (Rice et al., 1984), are not observed in the case of LAV. These are: a region 3' of the envelope gene consisting of a noncoding stretch (600–900 bp), followed by a coding sequence of 307–357 codons (X open reading frame), which may slightly overlap the U3 region of the LTR (Seiki et al., 1983; Rice et al., 1984; Sagata et al., 1984) and, second, the LTR being

Table 2. Comparison of the Size of the LAV LTR and LTR-Related Element to Those of Other Retroviruses

|         | LTR   | U3    | R   | US  | PU  | PBS | IR |
|---------|-------|-------|-----|-----|-----|-----|----|
| LAV     | 638   | 456   | 97  | 85  | 15  | LYS | 4  |
| HTLV-I  | 759   | 355   | 228 | 176 | 12' | PRO | 4' |
| HTLV-II | 763   | 314   | 249 | 261 | 12' | PRO | 4' |
| MMTV    | 1,332 | 1,197 | 11  | 124 | 19  | LYS | 8' |
| MoMuLV  | 594   | 449   | 68  | 77  | 13  | PRO | 13 |
| RSV     | 335   | 234   | 21  | 80  | 11  | TRP | 15 |
| SNV     | 601   | 420   | 97  | 80  | 13  | PRO | 9  |

Adapted from Chen and Barker (1984).

i = imperfect match or tract.

SNV = spleen necrosis virus (Shimotohno and Temin, 1982).

composed of unusually long US and R elements and the polyadenylation signal being situated in U3 instead of R (Seiki et al., 1983; Sagata et al., 1984; Shimotohno et al., 1984). We show here that, in contrast, the 3' end of the LAV envelope gene overlaps an open reading frame, termed F, that has the coding capacity for 206 amino acids and extends within the LTR (110 amino acids are encoded by the U3 region). The putatively encoded polypeptide (pF), the primary structure of which can be deduced, does not show any homology with the theoretical X gene products of the HTLV/BLV family. Also, the US and R elements are shorter (Table 2) and the polyadenylation signal is located within R, as is the case for all retroviruses except the HTLV/BLV. Additionally, LAV uses tRNA<sup>Gln</sup> as (-) strand primer, as opposed to tRNA<sup>Pro</sup> employed by all other mammalian retroviruses except MMTV (Donehower et al., 1981). Those homologies detected between the polymerase and protease domains of LAV and HTLV are also found in several retroviruses, RSV in particular.

It has been reported that a cloned HTLV-III genome hybridizes ( $T_m = 28^\circ\text{C}$ ) to sequences in the gag-pol and X regions of HTLV-I and -II; although restriction maps of cloned LAV and HTLV-III show almost perfect agreement (Hahn et al., 1984), we were unable to detect any such hybridization between LAV and HTLV-II ( $T_m = 55^\circ\text{C}$ ) (Alizon et al., 1984). Indeed, there is a punctual region of homology between LAV and HTLV-I (23/27 nucleotides starting at position 1859 in the LAV sequence) but nothing significant between the two viruses in the X region of HTLV-I. One possible reason for this discrepancy is that HTLV-III is subtly different from LAV. However it was subsequently reported that there was very minimal, if any, homology between ori X (of HTLV-I) and HTLV-III (Shaw et al., 1984).

#### Discussion

Regulatory sequences carried by retroviral LTR are believed to be involved in specific interactions between the viral genome and the host cell (Srinivasan et al., 1984). The LTR sequences of LAV are unique among retroviruses. That could reflect an original mode of gene expression, possibly in relation to particular transcriptional factors present in the virus-harboring cell. This hypothesis can be tested by studying the regulatory activity of the LAV

LTR sequences in transient or long-term experiments involving an indicator gene and different cellular contexts.

The presence of the Q and F reading frames in addition to the conventional gag-pol-env set of genes is unexpected. One should now address the question of their role in the viral cycle and pathogenicity by trying to characterize their protein product(s). It is tempting to speculate on a role of such polypeptide(s) in T4 cells' mortality, a problem that can be studied by designing synthetic peptides for antibody production or by using site-directed mutagenesis of Q and F coding regions.

The peculiar genetic structure of LAV poses the question of its origin. The virus shares common tracts with other (apparently unrelated) retroviruses. For instance, the unusually large size of the outer membrane glycoprotein (env) and a comparably sized genome are also observed in the case of lentiviruses such as Visna (Harris et al., 1981; Querat et al., 1984). The presence of a large part of the F open reading frame in the LTR, and the use of tRNA<sup>Gln</sup> as a primer for (-) strand synthesis, is reminiscent of the mouse mammary tumor virus. On the other hand, homologies in the pol gene would suggest that the LAV is closer to RSV than to any other retroviruses. Obviously, no clear picture can be drawn from the DNA sequence analysis as far as phylogeny is concerned. Thus, it may well be that LAV defines a new group of retroviruses that have been independently evolving for a considerable period of time, and not simply a variant recently derived from a characterized viral family. Both epidemiology and pathogeny of AIDS should be reconsidered with this idea in mind, when trying to answer such questions as these: Are there other human or animal diseases that are associated with similarly organized viruses? Is there a precursor to AIDS-associated virus(es) normally present, in latent form, in human populations? What triggered in this case the recent spreading of pathogenic derivatives?

#### Experimental Procedures

##### M13 Cloning and Sequencing

Total LAV DNA was sonicated, treated with the Klenow fragment of DNA polymerase plus deoxyribonucleotides (2 hr, 16°C), and fractionated by agarose gel electrophoresis. Fragments of 300–600 bp were excised, electroeluted, and purified by Elutip (Schleicher and Schüll) chromatography. DNA was ethanol-precipitated using 10 µg dextran T40 (Pharmacia) as carrier and ligated to dephosphorylated, Sma I-cleaved M13mp8 RF DNA using T4 DNA and RNA ligases (16 hr, 16°C) and transfected into *E. coli* strain TG-1. Recombinant clones were detected by plaque hybridization using the appropriate <sup>32</sup>P-labeled LAV restriction fragments as probes. Single-stranded templates were prepared from plaques exhibiting positive hybridization signals and were sequenced by the dideoxy chain termination procedure (Sanger et al., 1977) using  $\alpha$ -<sup>32</sup>P-dATP (Amersham, 400 Ci/mmol) and buffer gradient gels (Biggin et al., 1983). Sequences were compiled and analyzed using the programs of Staden adapted by B. Caudron for the Institut Pasteur Computer Center (Staden, 1982).

##### Strong-Stop cDNA

LAV virions from infected T lymphocyte (Barré-Sinoussi et al., 1983) culture supernatant were pelleted through a 20% sucrose cushion and the cDNA (-) strand was synthesized as described previously (Alizon et al., 1984) except that no exogenous primer was used. After alkaline hydrolysis (0.3 M NaOH, 30 min, 65°C), neutralization, and phenol extraction, the cDNA was ethanol-precipitated and loaded onto a 6%

acrylamide/8 M urea sequencing gel with sequence ladders as size markers.

#### Acknowledgments

We would like to thank Professors Luc Montagnier and Pierre Tiollais, in whose laboratory this work was carried out, for support and encouragement, as well as Professor Raymond Dedonder and Agnes Ullmann for their commitment to the project. Bernard Caudron and Jean-Noël Paulus of the Institut Pasteur Computer Center provided invaluable and constant assistance, and Michelle Fonck, technical support. Ana Cova and Louise-Marie Da Krelessy and good-humoredly typed the manuscript. We would like to thank Dr. Moshe Yaniv for critical reading of the manuscript and, finally, Genetic Systems, Seattle, WA, for communicating unpublished data.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 26, 1984

#### References

- Alizon, M., Sonigo, P., Barré-Sinoussi, F., Chermann, J. C., Tiollais, P., Montagnier, L., and Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. *Nature*, in press.
- Arya, S. K., Gallo, R. C., Hahn, B. H., Shaw, G. M., Popovic, M., Salaluddin, S. Z., and Wong-Staal, F. (1984). Homology of genome of AIDS-associated virus with genomes of human T-cell leukemia lymphoma viruses. *Science* 225, 927-930.
- Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Guillet, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome (AIDS). *Science* 220, 868-870.
- Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and <sup>32</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.
- Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucl. Acids Res.* 8, 1499-1504.
- Brun-Vézinet, F., Rouzioux, C., Barré-Sinoussi, F., Klatzmann, D., Saïmot, A. G., Rozenbaum, W., Montagnier, L., and Chermann, J. C. (1984). Detection of IgG antibodies to lymphadenopathy associated virus (LAV) by ELISA, in patients with acquired immunodeficiency syndrome or lymphadenopathy syndrome. *Lancet* I, 1253-1256.
- Chen, H. R., and Barker, W. C. (1984). Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. *Nucl. Acids Res.* 12, 1767-1778.
- Chen, I. S. Y., McLaughlin, J., Gasson, J. C., Clark, S. C., and Goldie, D. W. (1983). Molecular characterization of the genome of a novel human T-cell leukaemia virus. *Nature* 305, 502-505.
- Chiu, I. M., Callahan, R., Tronick, S. R., Scholm, J., and Aaronson, S. A. (1984). Major pol gene progenitors in the evolution of oncoviruses. *Science* 223, 364-370.
- Cianciolo, G. J., Kipnis, R. J., and Snyderman, R. (1984). Similarity between p15E of murine and feline viruses and p21 of HTLV. *Nature* 311, 515.
- Daly, H. M., and Scott, G. L. (1983). Fatal AIDS in a haemophiliac in the U.K. *Lancet* II, 1190.
- Ditmar, K. J., and Moelling, K. (1978). Biochemical properties of p15-associated protease in an avian RNA tumor virus. *J. Virol.* 28, 106-118.
- Donehower, L. A., Huang, A. L., and Hager, G. L. (1981). Regulatory and coding potential of the mouse mammary tumour virus long terminal redundancy. *J. Virol.* 37, 226-238.
- Gottlieb, M. S., Schroff, R., Schanler, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immuno-deficiency. *N. Eng. J. Med.* 305, 1426-1431.
- Hahn, B. H., Shaw, G. M., Arya, S. U., Popovic, M., Gallo, R. C., and Wong-Staal, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature* 312, 166-169.
- Harris, J. D., Scott, J. V., Taylor, B., Brahim, M., Stowring, L., Ventura, P., Haase, A. T., and Peluso, R. (1981). Viana virus DNA: discovery of a novel gapped structure. *Virology* 113, 573-583.
- Kiyokawa, T., Yoshikura, H., Hattori, S., Seiki, M., and Yoshida, M. (1984). Envelope proteins of human T-cell leukemia virus: expression in *Escherichia coli* and its application to studies of env gene functions. *Proc. Natl. Acad. Sci. USA* 81, 6202-6206.
- Klatzmann, D., Barré-Sinoussi, F., Nugeyre, M. T., Dauguet, C., Vilmer, E., Griscelli, C., Brun-Vézinet, F., Rouzioux, C., Gluckman, J. C., Chermann, J. C., and Montagnier, L. (1984). Selective tropism of lymphadenopathy-associated virus (LAV) for helper-inducer T-lymphocytes. *Science* 225, 59-63.
- Kozak, M. (1984). Compilation and analysis of sequences upstream from the transcriptional start site in eucaryotic mRNAs. *Nucl. Acids Res.* 12, 857-872.
- Lavy, J. A., Hoffman, A. D., Kramer, S. M., Lanois, J. A., Shimabukuro, J. M., and Oskro, L. S. (1984). Isolation of lymphocytotropic retroviruses from San Francisco patients with AIDS. *Science* 225, 840-842.
- Masur, H., Michelis, M. A., Greene, J. B., Onovato, I., Van de Stove, R. A., Holzman, R. S., Wormser, G., Bretman, L., Lange, M., Murray, H. W., and Cunningham-Rundles, S. (1981). An outbreak of community-acquired pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N. Eng. J. Med.* 305, 1431-1438.
- Misra, T. K., Grandgenett, D. P., and Parsons, J. T. (1982). Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. *J. Virol.* 44, 330-343.
- Montagnier, L., Chermann, J. C., Barré-Sinoussi, F., Chamarel, S., Guillet, J., Nugeyre, M. T., Rey, F., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Saïmot, A. G., Rozenbaum, W., Gluckman, J. C., Klatzmann, D., Vilmer, E., Griscelli, C., Gazengel, C., and Brunet, J. B. (1984). A new human T-lymphotropic retrovirus: characterization and possible role in lymphadenopathy and acquired immune deficiency syndromes. In *Human T-Cell Leukemia/Lymphoma Virus*, R. C. Gallo, M. E. Essex, and L. Gross, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 363-370.
- Oroszlan, S., Copeland, T. D., Kalyanaraman, V. S., Samgadharan, M. G., Schultz, A. M., and Gallo, R. C. (1984). Chemical analysis of human T-cell leukaemia virus structural proteins. In *Human T-Cell Leukemia/Lymphoma Virus*, R. C. Gallo, M. E. Essex, and L. Gross, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 101-110.
- Piot, P., Quinn, T. C., Taetle, H., Feinsod, F. M., Minlangu, K. B., Wogin, O., Mbendi, N., Mazebo, P., Ndangi, K., Stevens, W., Kalamay, K., Mitchell, S., Brids, C., and McCormick, J. B. (1984). Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *Lancet* II, 65-69.
- Popovic, M., Samgadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 224, 497-500.
- Querat, G., Barban, N., Sauze, N., Filippi, P., Vigne, R., Russo, P., and Vitto, C. (1984). Highly lytic and persistent lentiviruses naturally present in sheep with progressive pneumonia are genetically distinct. *J. Virol.* 52, 672-679.
- Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., Rajbhandary, U. L., and Gross, M. J. (1979). Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. *Eur. J. Biochem.* 97, 305-318.
- Rice, N. R., Stephen, R. M., Couez, D., Deschamps, J., Kettmann, R., Burny, A., and Gilde, R. V. (1984). The nucleotide sequence of the env gene and post-env region of bovine leukemia virus. *Virology* 138, 82-93.
- Sagata, N., Yasunaga, T., Ogawa, Y., Tsuzuku-Kawamura, J., and Ikawa, Y. (1984). Bovine leukemia virus: unique structural features of its long terminal repeats and its evolutionary relationship to human T-cell leukemia virus. *Proc. Natl. Acad. Sci. USA* 81, 4741-4745.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing

- with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Schüpbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Samgadhan, M. G., and Gallo, R. C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science 224, 503-505.
- Schwarz, D. E., Tizard, R., and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell 32, 853-869.
- Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA 80, 3618-3622.
- Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotoropic) virus type III in the acquired immune deficiency syndrome. Science 236, 1165-1171.
- Shimotohno, K., and Temin, H. M. (1982). Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J. Virol. 47, 163-171.
- Shimotohno, K., Golde, D. M., Miwa, M., Sugimura, T., and Chen, I. S. Y. (1984). Nucleotide sequence analysis of the long terminal repeat of human T-cell leukemia virus type II. Proc. Natl. Acad. Sci. USA 81, 1079-1083.
- Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukemia virus. Nature 293, 543-548.
- Srinivasan, A., Reddy, E. P., Dunn, C. Y., and Aaronson, S. A. (1984). Molecular dissection of transcriptional control elements with the long terminal repeat of retrovirus. Science 232, 286-289.
- Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids Res. 10, 4731-4751.
- Temin, H. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. Cell 27, 1-3.
- Weinberg, R. A. (1982). Fewer and fewer oncogenes. Cell 30, 3-9.